

# Complexity in the Immune System

Michael W. Deem

Department of Bioengineering and Department of Physics & Astronomy  
Rice University, Houston, TX 77005-1892  
mwdeem@rice.edu

February 4, 2008

Keywords: original antigenic sin, immune system, influenza, complexity

## Abstract

The immune system is a real-time example of an evolving system that navigates the essentially infinite complexity of protein sequence space. How this system responds to disease and vaccination is discussed. Of particular focus is the case when vaccination leads to increased susceptibility to disease, a phenomenon termed original antigenic sin. A physical theory of protein evolution to explain limitations in the immune system response to vaccination and disease is discussed, and original antigenic sin is explained as stemming from localization of the immune system response in antibody sequence space. This localization is a result of the roughness in sequence space of the evolved antibody affinity constant for antigen and is observed for diseases with high year-to-year mutation rates, such as influenza.

# 1 Introduction

Our immune system protects us against death by infection (Perelson and Weisbuch, 1997). A major component of the immune system is generation of antibodies, protein molecules that bind specific antigens. To recognize invading pathogens, the immune system performs a search of the amino acid sequence space of possible antibodies. To find useful antibodies in the effectively infinite protein sequence space, the immune system has evolved a hierarchical strategy. The first step involves creating the DNA sequences for B cells that code for moderately effective antibodies through rearrangement of immune-system-specific gene fragments from the genome (Tonegawa, 1983; Ehlich and Kuppers, 1995). This process is called VDJ recombination. This combinatorial process can produce on the order of  $10^{14}$  different antibodies through recombination of pieces of antibodies. The second step, which occurs when a specific antigen invades our body, is somatic hypermutation. Somatic hypermutation is the process of mutation that occurs when the B cells that produce the antibodies divide and multiply. Only those B cells that produce antibodies that bind the antigen with higher affinity are propagated by this mutation and selection process, and another name for this process is affinity maturation. Somatic hypermutation is essentially a search of the amino acid sequence space at the level of individual point mutations (Griffiths et al., 1984; Neuberger and Milstein, 1995; Tomlinson et al., 1996).

Binding of an antibody to an antigen results, when the immune system is protective, in clearance of the pathogen from which the antigen derives. Typical binding constants between antibodies and antigens are on the order of  $10^4$  to  $10^6$  or  $10^7$  l/mol (Janeway et al., 2001), although values as large as  $10^{11}$  l/mol or greater can be produced in experiments (Midelfort et al., 1999). Binding of antibody to antigen can prevent progression of disease by three mechanisms. If many antibodies bind and cover an invading pathogen, then the pathogen is prevented from entering cells and so fails to cause infection. Binding of antibodies to a pathogen may also cause phagocytic cells to ingest the pathogen. Finally, binding of antibodies to a pathogen may activate other components of the immune system, such as the complement, which contains pathogen-destroying proteins that punch holes in the membrane of the pathogen.

The space of possible antibodies is extremely large. Considering just the variable part of the heavy chain, there are on the order of  $20^{100} \approx 10^{130}$  possibilities. The immune system has evolved a hierarchical strategy for

searching this vast space of possible antibodies (Janeway et al., 2001). The hierarchy is composed of two steps. The first step is VDJ recombination, in which the three pieces of the variable region are joined together. Given the many possibilities for each of these three regions encoded within the genome of an individual, there are on the order of  $10^{12}$  possibilities of these random antibodies, or naive sequences. Of these, only on the order of  $10^8$  are expressed in a human individual at one point in time. Not all of these naive sequences lead to useful antibodies. Those antibodies that bind proteins of the self would cause autoimmune disease, and the B cells that produce them are killed. The B cells that produce antibodies that do not bind any proteins are left to die. But the B cells that produce antibodies that bind antigen multiply and divide. As the B cells divide, the DNA that codes for the antibodies mutates at a high rate. The process, termed somatic hypermutation, can be viewed as a local optimization of the initial, naive guesses.

The consequence of an immune system response to antigen is the establishment of a state of memory (Gray, 1993). Immunological memory is the ability of the immune system to respond more rapidly and effectively to antigens that have been encountered previously. Specific memory is maintained in the DNA of long-lived memory B cells that can persist without residual antigen (Black and Rosen, 1962; Sprent, 1993).

The adaptive vertebrate immune system is a wonder of modern evolution. Under most circumstances, the dynamics of the immune system is well-matched to the dynamics of pathogen growth during a typical infection. Some pathogens, however, have evolved escape mechanisms that interact in subtle ways with the immune system dynamics. In addition, negative interactions between the immune system, which has evolved over 400 000 000 years, and vaccination, which has been practiced for only 200 years, are possible.

Although our immune system is highly effective, some limitations have been reported. The phenomenon known as “original antigenic sin” is the tendency for antibodies produced in response to exposure to influenza virus antigens to suppress the creation of new, different antibodies in response to exposure to different versions of the flu (Fazekas de St. Groth and Webster, 1966a; Fazekas de St. Groth and Webster, 1966b). Roughly speaking, the immune system responds only to the antigen fragments, or epitopes, that are in common with the original flu virus. As a result, individuals vaccinated against the flu may become *more* susceptible to infection by mutated strains of the flu than would individuals receiving no vaccination.

The phenomenon of original antigenic sin has since been observed in dengue fever, human immunodeficiency virus (HIV), and other viruses. Dengue fever is caused by four distinct viral strains (Rigau-Perez et al., 1998). Once the immune system has memorized a response to one type of dengue, secondary infection with a different type of dengue places patients at the largest risk for dengue hemorrhagic fever (Halstead et al., 1983; Halstead, 1989; Vaughn et al., 2000). Thus, vaccination with one type of dengue virus can actually *increase* the rate of dengue hemorrhagic fever due to suppression of an immune response to the three other types of dengue virus. Similarly, the rapid replication of HIV, coupled with a high mutation rate, leads to the generation of many variants of HIV, even within a single infected patient (Condra et al., 1995; Finzi and Siliciano, 1998; Nara et al., 1990). HIV is initially controlled by cytotoxic T lymphocytes (CTLs), but may subsequently escape control through mutation of the relevant T-cell epitope, since CTLs also preferentially respond against the initial rather than mutated epitope (Good et al., 1993; Klenerman and Zinkernagel, 1998). Original antigenic sin of the CTL response impedes clearance of new HIV variants and enhances the chance that the immune system will be unable to control the HIV infection. The details of how original antigenic sin works, even at a qualitative level, are unknown.

Another early example of a limitation in the immune response came from the observation by Hoskins *et al.* that repeated annual influenza vaccination gave worse protection than did single vaccination among English school boys (Hoskins et al., 1979). Later studies, such as that by Keitel *et al.* observed that repeated vaccination is more effective than is single vaccination (Keitel et al., 1997).

While the Hoskins and Keitel experiments seem inconsistent, this is not necessarily the case. It may be that these experiments were performed under different conditions. In other words, there may be an additional order parameter that one needs to consider in order to understand these two experiments. This is indeed the case. The crucial order parameter is the number of accumulated mutations in the influenza virus, or “antigenic distance” (Smith et al., 1999). As parameterized by the antigenic distance, repeated vaccination was found to have higher efficacy than the single vaccination when the distance between the vaccine and the influenza was small, whereas repeated vaccination was found to be less effective when the distance between the vaccine and the influenza was large. Hoskins’ and Keitel’s are, thus, not necessarily inconsistent. It is not known, however, why the efficacy of single and repeated vaccination should vary as was observed in the experiments.

In this article, I offer an explanation for the reported limitations in the immune system response using a model of protein evolution (Deem and Lee, 2003). The dynamics of affinity maturation is described by a search in antibody sequence space for increased binding constants between antibody and antigen. It is shown that an immune system response to an antigen generates localized memory B cell sequences. This set of localized sequences reduces the ability of the immune system to respond to subsequent exposures to different but related antigens. It is this competitive process between memory sequences and the VDJ recombinations of secondary exposure that is responsible for the reported limitations in the immune system.

## 2 The Random Energy Model

A random energy model is used to represent the interaction between the antibodies and the influenza proteins. This model captures the essence of the correlated ruggedness of the interaction energy in the variable space, the variables being the antibody amino acid sequences and the identity of the disease proteins, and the correlations being mainly due to the physical structure of the antibodies. The random energy model allows study of the sequence-level dynamics of the immune/antigen system, which would otherwise be an intractable problem at the atomic scale, with  $10^4$  atoms per antibody,  $10^8$  antibodies per individual,  $6 \times 10^9$  individuals, and many possible influenza strains. Use of random energy theory to treat correlations in otherwise intractable physical systems goes back at least to Bohr's random matrix theory for nuclear cross sections (Bohr, 1936) and has been used for quantum chaos, disordered mesoscopic systems, QCD, and quantum gravity (Guhr et al., 1998). Close to the present application is the study of spin glasses by random energy models (Sherrington and Kirkpatrick, 1975; Derrida, 1980), protein folding by coarse-grained models (Bryngelson and Wolynes, 1987; Gutin and Shakhnovich, 1993), and evolutionary systems by NK-type models (Bogarad and Deem, 1999; Derrida and Peliti, 1991; Weisbuch, 1990; Drossel, 2001).

In detail, the generalized NK model used considers three different kinds of interactions within an antibody: interaction within a subdomain ( $U^{\text{sd}}$ ), interactions between subdomains ( $U^{\text{sd-sd}}$ ), and direct binding interaction between antibody and antigen ( $U^{\text{c}}$ ). In the context of protein evolution, parameters of the model have been calibrated (Bogarad and Deem, 1999; Kauffman and Levin, 1987; Perelson and Macken, 1995). The energy func-

tion of a protein is given by

$$U = \sum_{i=1}^M U_{\alpha_i}^{\text{sd}} + \sum_{i>j=1}^M U_{ij}^{\text{sd-sd}} + \sum_{i=1}^P U_i^{\text{c}} , \quad (1)$$

where  $M$  is the number of antibody secondary structural subdomains, and  $P$  is the number of antibody amino acids contributing directly to the binding. The subdomain energy  $U^{\text{sd}}$  is

$$U_{\alpha_i}^{\text{sd}} = \frac{1}{\sqrt{M(N-K+1)}} \sum_{j=1}^{N-K+1} \sigma_{\alpha_i}(a_j, a_{j+1}, \dots, a_{j+K-1}), \quad (2)$$

where  $N$  is the number of amino acids in a subdomain, and  $K$  is the range of local interaction within a subdomain. All subdomains belong to one of  $L = 5$  different types (*e.g.*, helices, strands, loops, turns, and others). The quenched Gaussian random number  $\sigma_{\alpha_i}$  is different for each value of its argument for a given subdomain type,  $\alpha_i$ . All of the Gaussian  $\sigma$  values have zero mean and unit variance. The energy of interaction between secondary structures is

$$U_{ij}^{\text{sd-sd}} = \sqrt{\frac{2}{DM(M-1)}} \sum_{k=1}^D \sigma_{ij}^k(a_{j_1}^i, \dots, a_{j_{K/2}}^i; a_{j_{K/2+1}}^j, \dots, a_{j_K}^j) . \quad (3)$$

The number of interactions between secondary structures is set to  $D = 6$ . Here  $\sigma_{ij}^k$  and the interacting amino acids,  $j_1, \dots, j_K$ , are selected at random for each interaction  $(k, i, j)$ . The chemical binding energy of each antibody amino acid to the antigen is given by

$$U_i^{\text{c}} = \frac{1}{\sqrt{P}} \sigma_i(a_i) . \quad (4)$$

The contributing amino acid,  $i$ , and the unit-normal weight of the binding,  $\sigma_i$ , are chosen at random. Using experimental results,  $P = 5$  amino acids are taken to contribute directly to the binding event. Here only five chemically distinct amino acid classes (*e.g.*, negative, positive, polar, hydrophobic, and other) are considered since each different type of amino acid behaves as a completely different chemical entity within the random energy model.

The generalized NK model, while a simplified description of real proteins, captures much of the thermodynamics of protein folding and ligand binding.

In the model, a specific B cell repertoire is represented by a specific set of amino acid sequences. Moreover, a specific instance of the random parameters within the model represents a specific antigen. An immune response that finds a B cell that produces an antibody with high affinity constant to a specific antigen corresponds in the model to finding a sequence having a low energy for a specific parameter set.

The random character of the generalized NK model makes the energy rugged in antibody sequence space. The energy is, moreover, correlated by the local antibody structure ( $K = 4$ ), the secondary antibody structure ( $U^{\text{sd-sd}}$ ), and the interaction with the influenza proteins ( $U^c$ ). As the immune system explores the space of possible antibodies, localization is possible if the correlated ruggedness of the interaction energy is sufficiently great.

### 3 Parameters of the Immune System

Since the variable region in each light and heavy chain of an antibody is about 100 amino acids long, and most of the binding typically occurs in one of the two chains, a single sequence length of 100 is chosen (Edelman, 1970; Porter, 1973). The value  $M = 10$  is chosen since there are roughly 10 secondary structures in a typical antibody and thus choose  $N = 10$ . The immune system contains of the order of  $10^8$  B cells divided into different specificities ( $10^7$  for the mouse) (Klinman and Press, 1975), and the frequency of a specific B cell participating in the initial immune response is roughly 1 in  $10^5$  (Janeway et al., 2001). Hence,  $10^3$  sequences are used during an immune response.

The hierarchical strategy of the immune system is used to search the antibody sequence space for high affinity antibodies. Initial combination of optimized subdomains is followed by a point mutation and selection procedure (Bogarad and Deem, 1999). To mimic combinatorial joining of gene segments during B cell development, a naive B cell repertoire is produced by choosing each subdomain sequence from pools that have  $N_{\text{pool}}$  amino acid segments obtained by minimizing the appropriate  $U^{\text{sd}}$ . To fit the theoretical heavy-chain diversity of  $3 \times 10^{11}$  (Janeway et al., 2001),  $N_{\text{pool}} = 3$  sequences are chosen from among the top 300 sequences for each subdomain type.

Somatic hypermutation occurs at the rate of roughly one mutation per variable regions of light and heavy chains per cell division, which occurs every 6 to 8 hours during intense cell proliferation (French et al., 1989; Kocks and Rajewsky, 1989).

Hence, in the simulation, 0.5 point mutations are done per sequence, the best (highest affinity)  $x = 20\%$  sequences are kept, and then these are amplified back up to a total of  $10^3$  copies in one round, which corresponds to 1/3 day. That is, the probability of picking one of the, possibly mutated, sequences for the next round is

$$p_{\text{select}} = \begin{cases} 1/200, & U \leq U_{200} \\ 0, & U > U_{200} \end{cases}, \quad (5)$$

where  $U_{200}$  is the 200th best energy of the  $10^3$  sequences after the mutation events, and this equation is employed  $10^3$  times to select randomly the  $10^3$  sequences for the next round. Given a specific antigen, *i.e.* a specific set of interaction parameters, 30 rounds (10 days) of point mutation and selection are performed in one immune response. In this way, memory B cells for the antigen are generated.

The affinity constant is given as a function of energy,

$$K^{\text{eq}} = \exp(a - bU), \quad (6)$$

where  $a$  and  $b$  are determined by the dynamics of the mutation and selection process. Affinity constants resulting from VDJ recombination are roughly  $10^4$ , affinity constants after the first response of affinity maturation are roughly  $10^6$ , and affinity constants after a second response of affinity maturation are roughly  $10^7$  l/mol (Janeway et al., 2001) These values fix the selection strength,  $x = 20\%$ . By comparison to the dynamics of the model, the values  $a = -18.56$  and  $b = 1.67$  are obtained.

The memory B cells, key to immunological memory, give rapid and effective response to the same antigen due to their increased affinity for previously-encountered antigens. The focus is on the role of the memory cells in the immune response to different antigens. The “distance” between a first antigen and the second antigen is given by the probability,  $p$ , of changing parameters of interaction within subdomain ( $U^{\text{sd}}$ ), subdomain-subdomain ( $U^{\text{sd-sd}}$ ), and chemical binding ( $U^c$ ) terms. Within  $U^{\text{sd}}$ , only the subdomain type,  $\alpha_i$ , is changed, not the parameters  $\sigma_\alpha$ , which are probably fixed by structural biology and should be independent of the antigen. This means that in the model, the primary response is always from the same set of theoretically possible naive antibodies, analogous to the primary response of one individual with a fixed set of naive possibilities from his or her genome.



## 4 Immune System Dynamics

The number of memory and naive B cells that participate in the immune response to the second antigen is estimated by the ratio of the respective affinity constants. From the definition of the affinity constant,

$$K^{\text{eq}} = \frac{[\text{Antigen} : \text{Antibody}]}{[\text{Antigen}] [\text{Antibody}]} , \quad (7)$$

the binding probability is proportional to the affinity constant and to the concentration of antigen-specific antibody, which is  $10^2$  times greater for the memory sequences (Janeway et al., 2001). The  $10^2$  times great concentration of memory antibodies is significant, because it biases the immune response toward the memory. The average affinity for the second antigen of the  $10^3$  memory cells,  $K_m^{\text{eq}}$ , and that of the  $10^3$  B cells from the naive repertoire of optimized subdomain sequences,  $K_n^{\text{eq}}$ , are measured. The ratio  $10^2 K_m^{\text{eq}} / K_n^{\text{eq}}$  gives the ratio of memory cells to naive cells. For exposure to the second antigen, 30 rounds (10 days) of point mutation and selection are performed, starting with  $10^5 K_m^{\text{eq}} / (10^2 K_m^{\text{eq}} + K_n^{\text{eq}})$  memory cells and  $10^3 K_n^{\text{eq}} / (10^2 K_m^{\text{eq}} + K_n^{\text{eq}})$  naive cells, since both memory and naive sequences participate in the secondary response (Berek and Milstein, 1987).

Before going in to more details of the immune system response, I mention two possibilities for suboptimal dynamics. The first is localization in antibody amino-acid sequence space. Imagine the binding energy of an antibody for this year's flu as a function of antibody sequence. There will be a region where the best antibodies are located. The binding energy of antibodies for next year's flu will be slightly different, as the flu will likely have mutated during the year. The region where the best antibodies are to be found next year will be slightly different from the region for this year. There may still be a local minimum for the binding energy of next year's flu to antibodies around the best region for this year's flu. If this is the case, that means there is a barrier that separates the region of best antibodies from one year to the next. This, in turn, means the immune system must overcome this barrier in order to evolve the antibodies during the year. If, for example, the individual dies from the flu before this barrier is overcome, this really is localization of antibodies in sequence space. Note that the barrier may be both energetic and entropic. That is, the mutations required to move from one region to another may either lead to less favorable antibodies along the

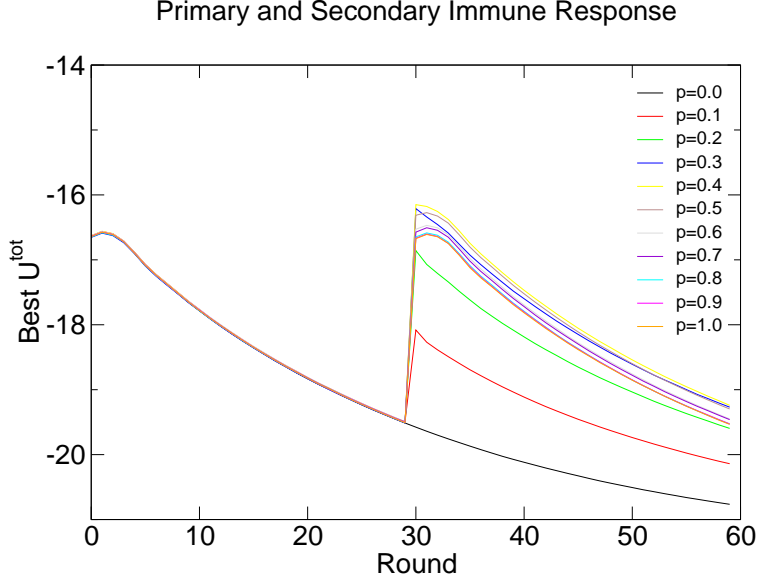


Figure 1: The dynamics of the immune response within the model. A vaccine is given in the first year. The individual is exposed to the disease in the second year. The disease differs from the vaccine by an amount  $p$ . The results, as with all presented here, are averaged over 5000 ensembles.

way, or they may be more numerous than would be expected simply based upon the amino acid differences between the two optimal regions

## 5 Original Antigenic Sin

Some specific results from the model clarify the concepts involved. In Fig. 1 is shown the dynamics of the immune response. Again, the starting point is from  $10^3$  naive sequences. These sequences are evolved for 30 rounds. At each round, the sequences are mutated at a rate of 0.5 mutations per sequence per round, and the top 20% are kept. After the vaccination in year 1, the person is exposed to the disease in year 2. The disease may be similar to or very different from the vaccine. The degree of similarity is measured by the parameter  $p$ . The secondary response starts with some naive sequences and some memory sequences, weighted by the average binding constants and the relative concentrations of each.

A key point about these results is that localization does occur. Shown in

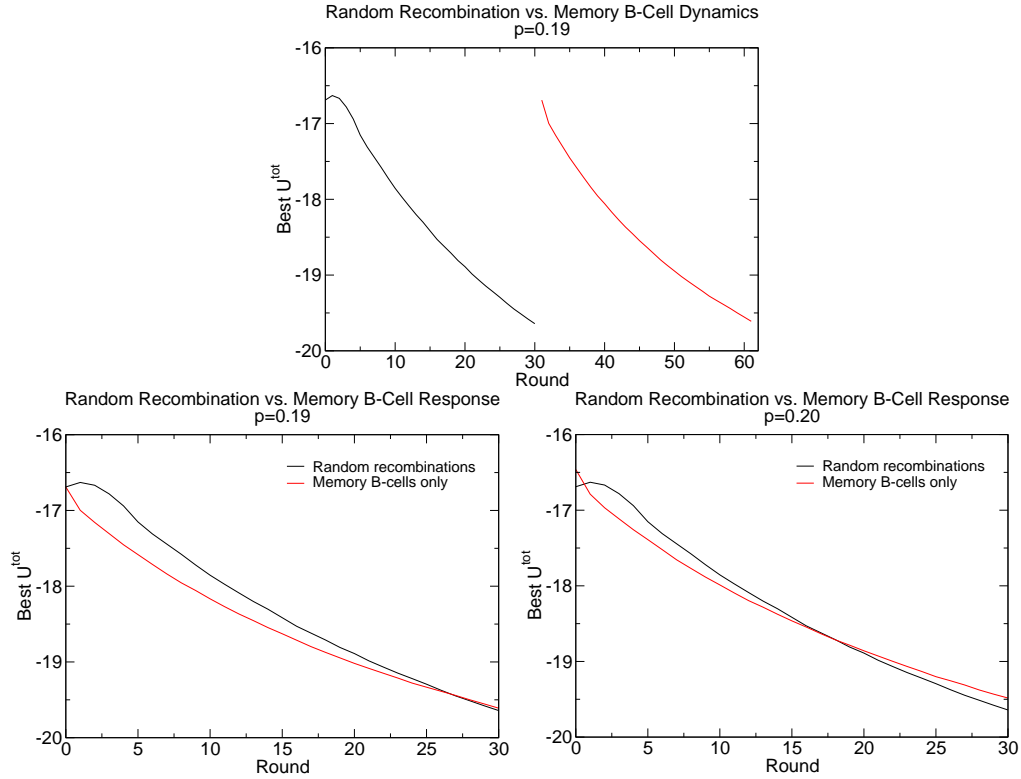


Figure 2: a) Comparison of primary and secondary response when the secondary response is from memory sequences only. b) Comparison when secondary response is shifted to the left to overlap with primary. c) Comparison for a larger value of  $p$ .

2 is the behavior of the the primary and secondary response. In this figure only, the secondary response is exclusively from memory sequences. It can be seen that in the early part of the secondary response, the memory sequences look superior to the naive sequences. By the end of 30 rounds, however, the memory sequences are worse than the naive sequences due to an inferior ability to evolve.

The results from 2 are somewhat academic, in that the true secondary response was not used. In figure 3 is shown the response when the true secondary response is used. It is again observed that the memory sequences look promising at the beginning of the response, but they are unable to evolve

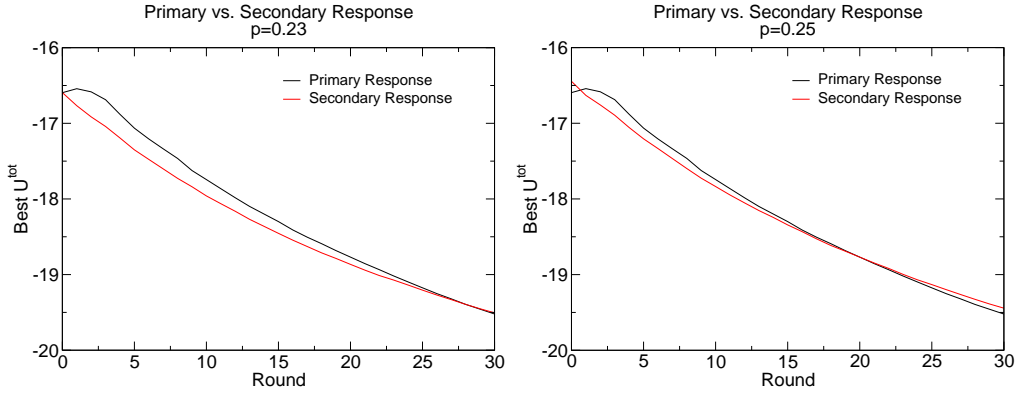


Figure 3: a) Comparison of primary and secondary response when the secondary response is from memory and naive sequences. b) Comparison for a larger value of  $p$ .

as well as naive sequences. In other words, the presence of the vaccine that generated the memory response leads to a worse response than if there were no vaccine. This is a demonstration of original antigenic sin.

It is of interest to measure the degree of original antigenic sin as a function of the mutation of the flu. That is, one wants to compare the secondary response at day 10 with a primary response at day 10. Shown in Fig. 4 is this comparison. The secondary response is better than the primary when the vaccine and flu are similar. When the vaccine and flu are very different, the secondary response uses mostly naive sequences, and it becomes like a primary response. In between, there is the region of original antigenic sin.

It is of interest to measure the degree of original antigenic sin as a function of the mutation of the flu. That is, Fig. 5 shows the evolved affinity constant to a second antigen if the exposure to a first antigen exists (solid line) or not (dashed line) as a function of the difference between the first and second antigen,  $p$ , or “antigenic distance” (Smith et al., 1999). When the difference is small, the exposure to a first antigen leads to higher affinity constant than without exposure, which is why immune system memory and vaccination is effective. For a large difference, the antigen encountered in the first exposure is uncorrelated with that in the second exposure, and so immune system memory does not play a role. Interestingly, the immunological memory from the first exposure actually gives worse protection, *i.e.* a lower

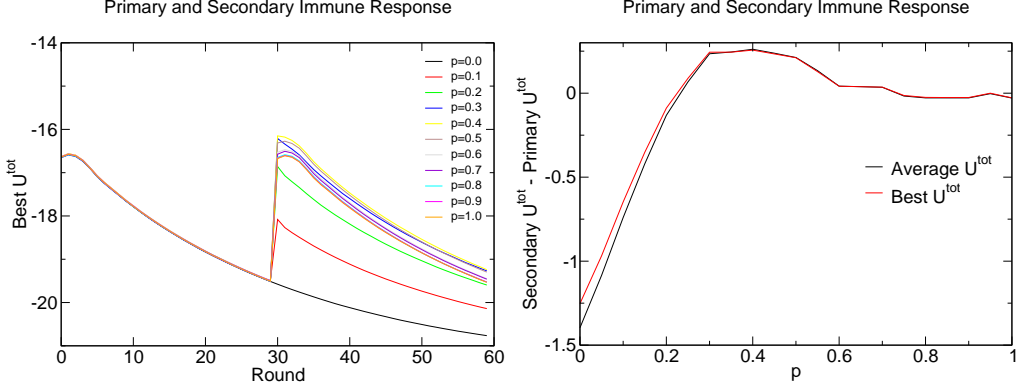


Figure 4: a) The model dynamics. The effectiveness of vaccination is determined by comparing the secondary response at round 60 with the primary response at round 30. b) The energy evolved in the secondary response minus that evolved in the primary response.

affinity constant, for intermediate differences—which is original antigenic sin.

The dynamics of the immune response (Fig. 4) depend upon the constants of Nature, *i.e.* the parameters of the model. For example, in an organism with a smaller immune system, such as the mouse, there are fewer starting sequences, and less favorable binding constants are measured in the same number of rounds: A factor of 0.5 reduction in the number of starting sequences leads to a 0.64 reduction in the evolved binding constant, but a similar degree of original antigenic sin as in Fig. 4, as shown in Fig. 6. The original antigenic sin curve looks very similar for the two repertoire sizes. The only significant difference is that the absolute values of the evolved energies are slightly worse for the smaller repertoire size, due to the reduction in diversity.

On the other hand, if more rounds are performed, better binding constants are found in the primary and secondary responses, but the secondary response is not as improved over the primary response as when using 30 rounds. This is because the evolved sequences are becoming more localized in ever deeper wells. The degree of original antigenic sin is, however, similar in the range of 30 to 60 rounds per response. The model dynamics are shown in Fig. 7. While 60 rounds is really too many for a single response,

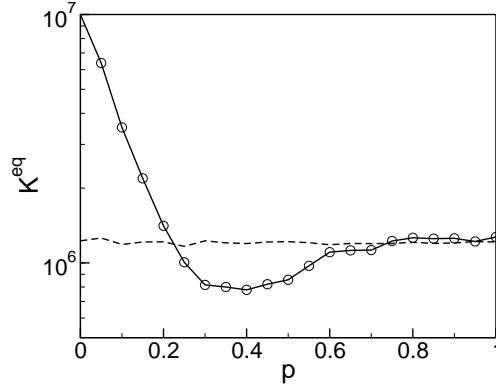


Figure 5: The evolved affinity constant to a second antigen after exposure to an original antigen that differs by probability  $p$  (solid line). The dotted line represents the affinity constant without previous exposure. The affinity constant is generated by exponentiating, as in Eq. (6), the average of the best binding energy.

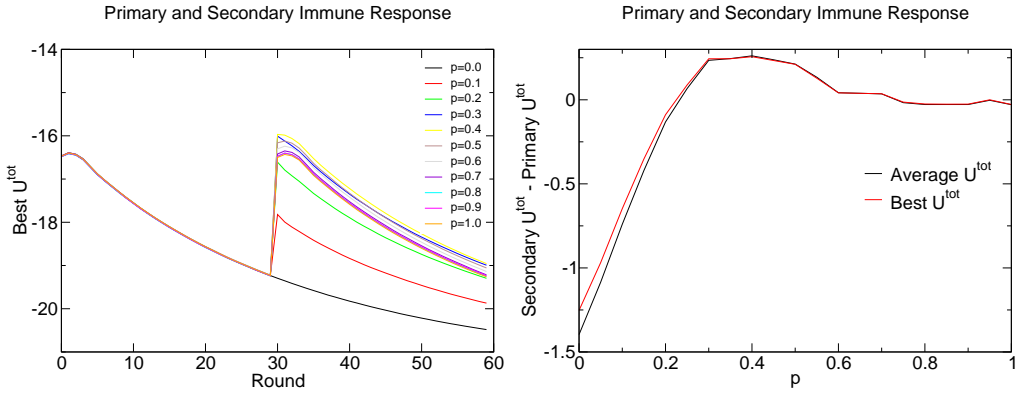


Figure 6: a) The model dynamics with a repertoire size of 500 instead of 1000. The effectiveness of vaccination is determined by comparing the secondary response at round 60 with the primary response at round 30. b) The energy evolved in the secondary response minus that evolved in the primary response.

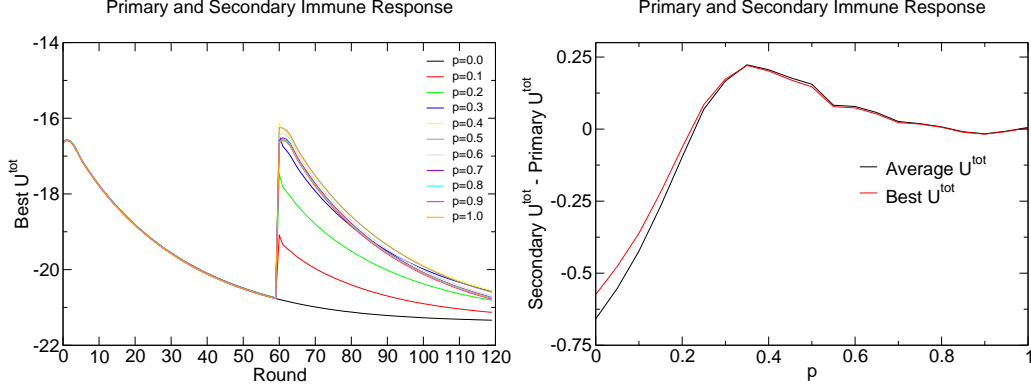


Figure 7: a) The model dynamics with 60 rounds instead of 30. The effectiveness of vaccination is determined by comparing the secondary response at round 120 with the primary response at round 60. b) The energy evolved in the secondary response minus that evolved in the primary response.

this experiment could represent administration of a vaccine twice in year one, administration of a vaccine once in year two, and exposure to the disease in year two. In this case, one imagines that the two vaccines in year one are identical and that the vaccine and disease in year two are identical.

Similarly, if the roughness of the energy upon disease mutation is increased, for example by assuming that mutation of the influenza actually changes the  $\sigma_\alpha$ , the degree of original antigenic sin increases substantially, by a factor of 2, because the barriers between the regions of localization in sequence space are increased. On the other hand, if the concentration of the memory cells is decreased, the contribution of the memory cells to the dynamics is reduced, and the original antigenic sin phenomenon decreases, almost disappearing when the memory and naive antibody concentrations are equal. Larger selection strengths ( $x < 20\%$ ) cause more localization and original antigenic sin, in shallower wells for small  $x$ , and smaller strengths ( $x > 20\%$ ) lead to less evolution.

The average number of mutations leading to the best antibody, a measure of the localization length when original antigenic sin occurs, is 15 for the first response and rises from 5 to 15 for the second response in the range  $0 \leq p < 0.30$ , as shown in Fig. 8. Interestingly, the average number of

Mutational Difference between Primary and Secondary Response

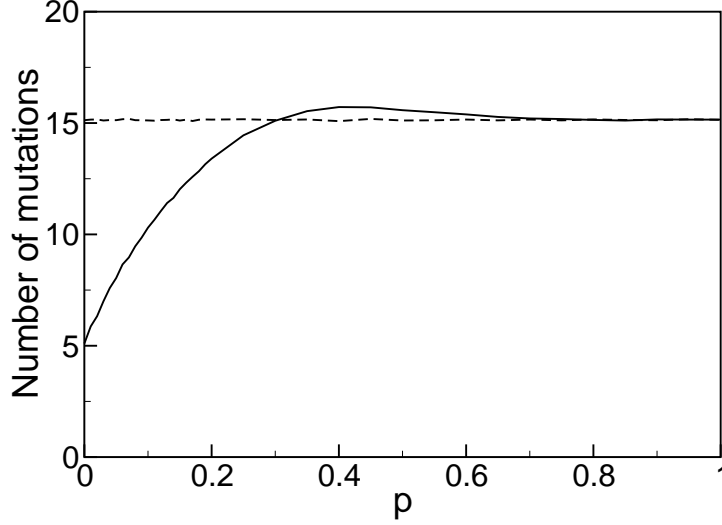


Figure 8: The number of mutations that occur in the best evolved antibody during the primary (dashed) and secondary (solid) responses.

mutations rises slightly above 15 in the range  $0.30 \leq p < 0.70$ , indicating that in the original antigenic sin region more mutations are necessary for the compressed ensemble of memory sequences from the primary response to evolve to a suitable new state in the secondary response. This result implies that there is at least some entropic character to the barrier between the optimal regions of antibody sequence space from one year to the next.

For small values of  $p$ ,  $p < 0.19$ , the memory B cells produce antibodies with higher affinities,  $K_m^{\text{eq}} > 10^4$  l/mol, for the new antigen than do naive B cells. The binding constant of the memory antibodies steadily decreases with  $p$ , reaching the non-specific value of  $K_m^{\text{eq}} = 10^2$  at  $p = 0.36$ , as shown in Fig. 9. Interestingly, this is less than the range to which original antigenic sin extends,  $0.23 < p < 0.60$ . These model predictions are in good agreement with experimental data on cross-reactivity, which ceases to occur when the amino acid sequences are more than 33–42% different (East et al., 1980). This comparison also shows that the parameter  $p$ , which in the model is the probability with which the parameters in the random energy model are



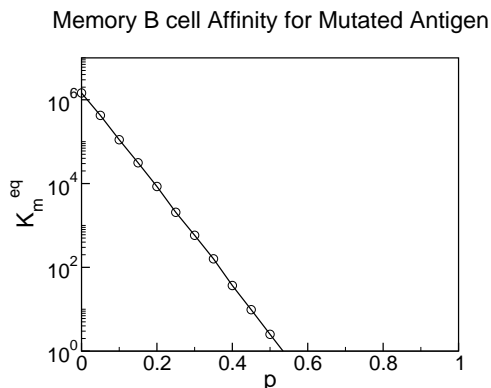


Figure 9: The affinity constant of an antibody as the antigen to which it evolved against is mutated.

changed, can be interpreted as the degree of amino acid difference.

The ineffectiveness of immune system memory over a window of  $p$  values can be understood from the localization of memory B cell sequences. Figure 10 displays distributions of memory and naive affinity constants for the second antigen. Notice that the memory sequences are highly homogeneous and lack diversity compared to the naive sequences. Indeed, original antigenic sin arises mainly because the memory sequences from the primary response suppress use of naive sequences in the tail of the distribution for the secondary response. Although those naive sequences initially look unpromising, they may actually evolve to sequences with superior binding constants. Figure 11 illustrates this phenomenon. Interestingly, when half of the distribution is removed, the reduction in the binding constant is just about that which occurs in original antigenic sin, Fig. 5.

The phenomenon of original antigenic sin, then, arises because the memory sequences cut off the tail of the naive sequences with poor binding constants. In most cases, *i.e.* for most values of  $p$ , this is an appropriate strategy. In some cases, however, those unpromising looking sequences actually lead to a better evolved binding constant than do the memory sequences. It is these cases in which original antigenic sin occurs. This mechanism can be tested by allowing evolution of only the top fraction of naive sequences. That is, the top fraction of the sequences is kept, overwriting the bottom fraction with

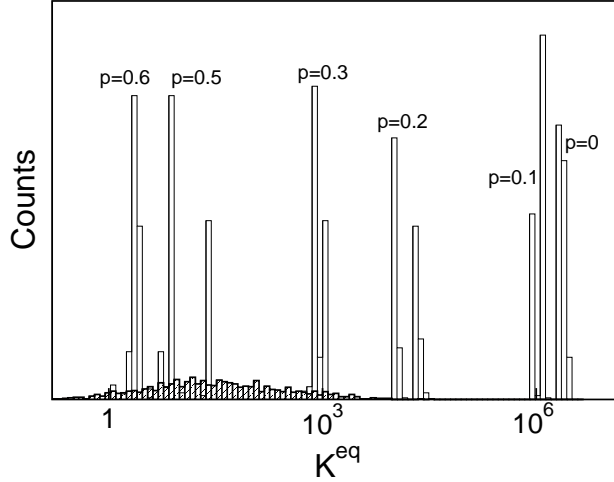


Figure 10: The affinity distribution of antibodies from naive B cells (hatched) or memory cells (open) for the antigen of second exposure.

the worst kept sequence. The repertoire size is still  $10^3$ , but the diversity is the fraction kept times  $10^3$ . Shown in Fig. 11 is the result. It is seen that if only the best sequences is kept, a very poor evolved antibody results. If only the 50% best sequences are kept, the best evolved antibody is worse by just about the depth of the original antigenic sin valley in Fig. 5.

## 6 Single and Repeated Vaccination

The model is also used to address the effectiveness of repeated annual influenza vaccination. Repeated vaccination will lead to a more effective response against closely related antigens. The ratio of the affinity constant to influenza after repeated vaccination,  $K_r^{\text{eq}}$ , to that after single vaccination,  $K_s^{\text{eq}}$ , should be greater than unity for small  $p$  and converge to unity for  $p \rightarrow 1$ . Repeated vaccination generates highly localized memory cells that have higher affinity constants to both the vaccine and the influenza if  $p$  is small. Results for our model are shown in Figs. 12 and 13. The decay of the ratio  $K_r^{\text{eq}}/K_s^{\text{eq}}$  to unity is in accord with experimental data (Smith et al., 1999). Interestingly, for our present vaccination protocol, no localization is observed,

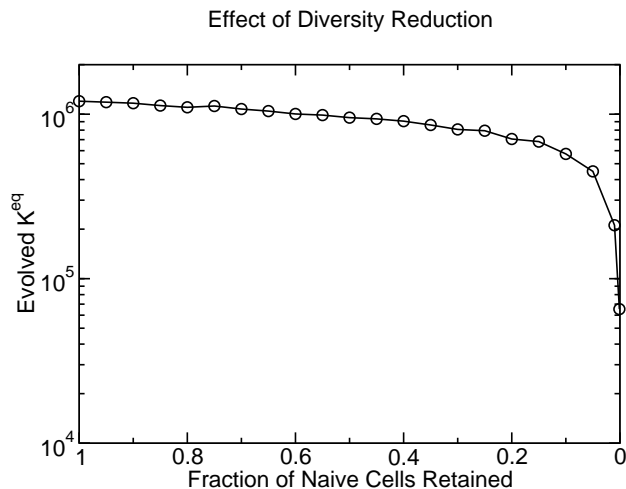


Figure 11: The evolved binding constant when only the top fraction of naive sequences are kept in the dynamics.

which is consistent with most of the modern experiments and with current public health policy that recommends for repeated, annual vaccinations of individuals at risk.

## 7 Summary

In summary, the generalized NK model is shown to successfully model immune system dynamics. A localization mechanism for the original antigenic sin phenomenon observed in the flu is explained. Localization of antibodies in the amino acid sequence space around memory B cell sequences is shown to lead to a decreased ability of the immune system to respond to diseases with year-to-year mutation rates within a critical window. This localization occurs because of the ruggedness of the evolved affinity constant in amino acid sequence space.

From the model dynamics, it is found that memory sequences can both out compete the non-vaccinated immune response and become trapped in local minima. Memory sequences with affinity constants initially superior to those from naive sequences can be selected by the dynamics, and these mem-

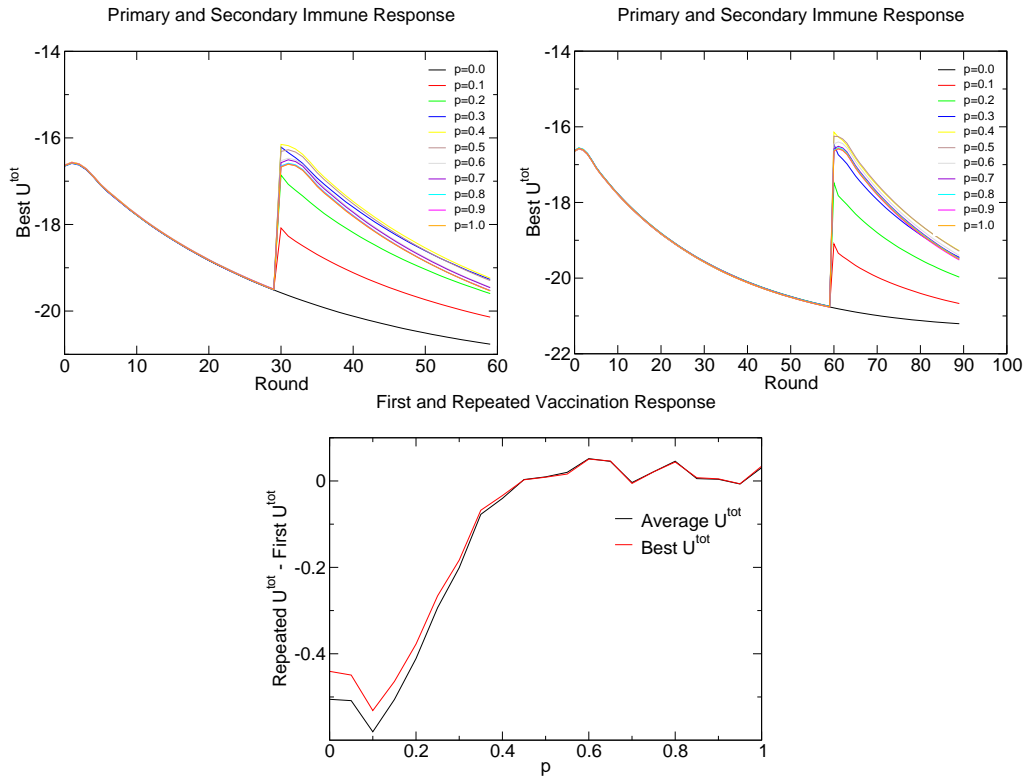


Figure 12: a) Dynamics with single vaccination. b) Dynamics with repeated vaccination. c) Comparison between single and repeated vaccination.

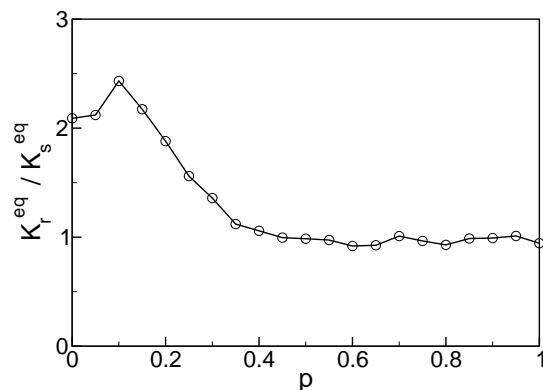


Figure 13: The ratio of the affinity constants to influenza after repeated vaccination and single vaccination.

ory sequences can lead to poorer evolved affinity constants, to the detriment of the immune system for intermediate disease mutation rates. Interestingly, due to the high diversity of the naive sequences, the localization phenomenon extends even beyond the region of cross-reactivity between memory antibodies and mutated antigen.

These results suggest several implications for vaccination strategy: the difference between vaccinations administered on a repeated basis should be as great as practicable, and suppression of the memory B cell response may be helpful during vaccination against highly variable antigens.

## Acknowledgements

The author thanks Jeong-Man Park for stimulating discussions. This research was supported by the National Science and Camille & Henry Dreyfus Foundations.

## References

Berek, C. and Milstein, C. (1987). Mutation drift and repertoire shift in the maturation of the immune response. *Immuno. Rev.*, 96:23–41.

- Black, F. L. and Rosen, L. (1962). Patterns of measles antibodies in residents of Tahiti and their stability in the absence of re-exposure. *J. Immunol.*, 88:725–731.
- Bogard, L. D. and Deem, M. W. (1999). A hierarchical approach to protein molecular evolution. *Proc. Natl. Acad. Sci. USA*, 96:2591–2595.
- Bohr, N. (1936). Neutron capture and nuclear constitution. *Nature*, 137:344–348.
- Bryngelson, J. D. and Wolynes, P. G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA*, 84:7524–7528.
- Condra, J. H., Schleif, W. A., Blahy, O. M., Gabryelski, L. J., Graham, D. J., Quintero, J. C., Rhodes, A., Robbins, H. L., Roth, E., Shivaprakash, M., Titus, D., Yang, T., Teppler, H., Squires, K. E., Deutsch, P. J., and Emini, E. A. (1995). In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature*, 374:569–571.
- Deem, M. W. and Lee, H.-Y. (2003). Sequence space localization in the immune system response to vaccination and disease. *Phys. Rev. Lett.*, 91:068101—1–4.
- Derrida, B. (1980). Random energy model—Limit of a family of disordered models. *Phys. Rev. Lett.*, 45:79–82.
- Derrida, B. and Peliti, L. (1991). Evolution in a flat fitness landscape. *Bull. Math. Biol.*, 53:355–382.
- Drossel, B. (2001). Biological evolution and statistical physics. *Adv. Phys.*, 50:209–295.
- East, J. J., Todd, P. E. E., and Leach, S. J. (1980). Original antigenic sin: Experiments with a defined antigen. *Mol. Immunol.*, 17:1545–1551.
- Edelman, G. M. (1970). The structure and function of antibodies. *Sci. Amer.*, 223:34–42.
- Ehlich, A. and Kuppers, R. (1995). Analysis of immunoglobulin gene rearrangements in single B-cells. *Curr. Opin. Immunol.*, 7:281–284.

- Fazekas de St. Groth, S. and Webster, R. G. (1966a). Disquisition on original antigenic sin: Evidence in man. *J. Exp. Med.*, 124:331–345.
- Fazekas de St. Groth, S. and Webster, R. G. (1966b). Disquisition on original antigenic sin: Proof in lower creatures. *J. Exp. Med.*, 124:347–361.
- Finzi, D. and Silliciano, R. F. (1998). Viral dynamics in HIV-1 infection. *Cell*, 93:665–671.
- French, D. L., Laskov, R., and Scharff, M. D. (1989). The role of somatic hypermutation in the generation of antibody diversity. *Science*, 244:1152–1157.
- Good, M. F., Zevering, Y., Currier, J., and Bilsborough, J. (1993). Original antigenic sin, T cell memory and malaria sporozoite immunity: An hypothesis for immune evasion. *Parasite Immunol.*, 15:187–193.
- Gray, D. (1993). Immunological memory. *Annu. Rev. Immun.*, 11:49–77.
- Griffiths, G. M., Berek, C., Kaartinen, M., and Milstein, C. (1984). Somatic mutation and the maturation of immune-response to 2-phenyl oxazolone. *Nature*, 312:271–275.
- Guhr, T., Müller-Groeling, A., and Weidenmüller, H. A. (1998). Random-matrix theories in quantum physics: Common concepts. *Phys. Rep.*, 299:189–425.
- Gutin, A. M. and Shakhnovich, E. I. (1993). Ground state of random copolymers and the discrete random energy model. *J. Chem. Phys.*, 98:8174–8177.
- Halstead, S. B. (1989). Antibody, microphages, dengue virus-infection, shock, and hemorrhage—A pathogenetic cascade. *Rev. Infect. Dis.*, 11 Supple. 4:S830–S839.
- Halstead, S. B., Rojanasuphot, S., and Sangkawibha, N. (1983). Original antigenic sin in dengue. *Am. J. Hyg. Trop. Med.*, 32:154–156.
- Hoskins, T. W., Davis, J. R., Smith, A. J., Miller, C. L., and Allchin, A. (1979). Assessment of inactivated influenza-A vaccine after 3 outbreaks of influenza-A at Christs Hospital. *Lancet*, i:33–35.

- Janeway, C. A., Travers, P., Walport, M., and Shlomchik, M. (2001). *Immunobiology*. Garland Publishing, New York, 5th edition.
- Kauffman, S. and Levin, S. (1987). Towards a general-theory of adaptive walks on rugged landscapes. *J. Theor. Biol.*, 128:11–45.
- Keitel, W. A., Cate, T. R., Couch, R. B., Huggins, L. L., and Hess, K. R. (1997). Efficacy of repeated annual immunization with inactivated influenza virus vaccines over a five year period. *Vaccine*, 15:1114–1122.
- Klenerman, P. and Zinkernagel, M. (1998). Original antigenic sin impairs cytotoxic T lymphocyte responses to viruses bearing variant epitopes. *Nature*, 394:482–485.
- Klinman, N. R. and Press, J. L. (1975). B cell specificity repertoire: Its relationship to definable subpopulations. *Transplant. Rev.*, 24:41–83.
- Kocks, C. and Rajewsky, K. (1989). Stable expression and somatic hypermutation of antibody V regions in B-cell developmental pathways. *Annu. Rev. Immunol.*, 7:537–559.
- Midelfort, K. S., Boder, E. T., and Wittrup, K. D. (1999). Dissecting the pathway of affinity maturation of a femtomolar anti-fluorescein single chain antibody fragment. *Biophys. J.*, 76:A111.
- Nara, P. L., Smit, L., Dunlop, N., Hatch, W., Merges, M., Waters, D., Kelliher, J., Gallo, R. C., Fischinger, P. J., and Goudsmit, J. (1990). Emergence of viruses resistant to neutralization by V3-specific antibodies in experimental human immunodeficiency virus type 1 IIIB infection of chimpanzees. *J. Virol.*, 64:3779–3791.
- Neuberger, M. S. and Milstein, C. (1995). Somatic hypermutation. *Curr. Opin. Immunol.*, 7:248–254.
- Perelson, A. S. and Macken, C. A. (1995). Protein evolution on partially correlated landscapes. *Proc. Natl. Acad. Sci. USA*, 92:9657–9661.
- Perelson, A. S. and Weisbuch, G. (1997). Immunology for physicists. *Rev. Mod. Phys.*, 69:1219–1267.
- Porter, R. R. (1973). Structural studies of immunoglobulins. *Science*, 180:713–716.



- Rigau-Perez, J. G., Clark, G. G., Gubler, D. J., Reiter, P., Sanders, R. J., and Vorndam, A. V. (1998). Dengue and dengue haemorrhagic fever. *Lancet*, 352:971–977.
- Sherrington, D. and Kirkpatrick, S. (1975). Solvable model of a spin glass. *Phys. Rev. Lett.*, 35:1792–1796.
- Smith, D. J., Forrest, S., Ackley, D. H., and Perelson, A. S. (1999). Variable efficacy of repeated annual influenza vaccination. *Proc. Natl. Acad. Sci. USA*, 96:14001–14006.
- Sprent, J. (1993). Life-spans of naive, memory and effector lymphocytes. *Curr. Opin. Immunol.*, 5:433–438.
- Tomlinson, I. M., Walter, G., Jones, P. T., Dear, P. H., Sonnhamme, E. L. L., and Winter, G. (1996). The imprint of somatic hypermutation on the repertoire of human germline V genes. *J. Mol. Biol.*, 256:813–817.
- Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature*, 302:575–581.
- Vaughn, D. W., Green, S., Kalayanarooj, S., Innis, B. L., Nimmannitya, S., Suntayakorn, S., Endy, T. P., Raengsakulrach, B., Rothman, A. L., Ennis, F. A., and Nisalak, A. (2000). Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. *J. Infec. Dis.*, 181:2–9.
- Weisbuch, G. (1990). A shape space approach to the dynamics of the immune system. *J. Theor. Biol.*, 143:507–522.